Who Are You, Elena Ferrante? A Computational Authorship Analysis via Clustering and Unmasking Methods

Anonymous ACL submission

Abstract

The aim of the current paper is to bring a contribution to the ongoing debate regarding the true identity of the author behind the pen name Elena Ferrante using stylometric methods that have not been applied to this problem before. The two main methods we used were hierarchical clustering based on rank distance and umasking. All experiments pointed towards Domenico Starnone as the author behind the Ferrante corpus, strengthening the conclusions of previous research.

1 Introduction and Related Work

Authorship attribution is a centuries-old dilemma, as works with disputed or anonymous authors date as far as the first encounters of written texts. Modern computational methods have made significant contributions to the field by quantitatively analyzing writing styles, linguistic patterns, and other features that are unique to individual authors (Stamatatos, 2009). Contemporary stylometry-based authorship attribution methods stem from the assumption that each writer will leave a unique and unconscious fingerprint on their work, comprising of lexical and syntactic patterns and which is more or less consistent throughout an entire body of works (Van Halteren et al., 2005; Koppel et al., 2009; Holmes, 1994; Juola, 2008).

While the field of stylistics in literary theory aims to capture and interpret the purposeful syntactic and idiomatic choices of the author alongside the meanings they convey (Shen, 2017), authorship computational analyses are concerned with the so-called "human stylome" (Van Halteren et al., 2005) - the unconscious language patterns. We thus make a distinction between *style*, which entails a certain degree of intention, and *stylistic fingerprint*, which is unique and unconscious for each individual writer.

Up to date, numerous authorship controversies have been investigated or solved by means of computational methods (Mosteller and Wallace, 1964; Craig and Kinney, 2009; Labbé and Labbé, 2001; Juola, 2015). In the last two decades, readers, journalists and computational linguists have been riddled by the case of Elena Ferrante, an anonymous author who has risen to international fame with the trilogy *L'amica geniale*, deemed the *New York Times'* No. 1 book of the century (The New York Times, 2024). The true identity of the writer remains however one of the best-kept editorial secrets, and numerous journalistic investigations haven't been able to uncover the truth. Several Italian authors, both male and female, have been proposed as the real writer behind the Ferrante novels.

In 2016, Italian journalist Claudio Gatti claimed to have uncovered the real "Ferrante" by tracing the payments made by her publishers all the way to Anita Raja, a translator from German and the wife of writer Domenico Starnone (Gatti, 2016). In 2017 however, University of Padua organized an invitation-based workshop (Tuzzi et al., 2018a) where international experts conducted a series of experiments based on a corpus of 150 novels written by 40 Italian authors. The collective consensus was that the Ferrante novels were most likely written by Domenico Starnone. The similarity between Doemico Starnone and Elena Ferrante's novels was found to be so high that Swiss researcher Jacque Savoy offered a reward of 20 Euros to whomever might bring compelling scientific proofs that the real author behind the novels is someone other than Starnone (Savoy, 2018b). Nevertheless, Starnone denied all claims.

Numerous computational methods have been employed in the case of Elena Ferrante, including approaches based on semantic similarity (Juola, 2017) and methods based on word frequencies such as Burrow's Delta and Labbé's distance (Tuzzi et al., 2018b; Savoy, 2018a; Rybicki, 2018). The cur-

rent paper aims to bring a new contribution to this remarkable debate by employing authorship attribution methods that so far have never been used on the Ferrante corpus. Taking advantage of the insights provided by previous research, our goal was to answer two main questions: (1) Can we further the claim that Domenico Starnone is the writer behind Elena Ferrante, and (2) Can we find evidence supporting the Anita Raja lead? To address these questions, the study is structured as follows: Section 2 outlines our methodology, followed by a description of the experiments and results in Section 3. Finally, Section 4 presents and analyses our conclusions.

2 Methodology

In order to investigate the authorship of Elena Ferrante's novels, a first step was to generate a corpus of works written by Elena Ferrante, as well as works of possible candidates. We followed the description of the Padova corpus and decided to select a smaller amount of authors, each represented of at least 4 works. For a detailed description of the pre-processing steps, see Section 2.1.

Before delving into the two main authorship attribution experiments, we visually inspected the data via PCA (Principal Component Analysis) in order to reveal some insights regarding potential similarities among authors.

Then, to test the authorship hypotheses, we employed two techniques based on supervised and unsupervised machine learning algorithms. The first approach was hierarchical clustering based on similarity distances. As a similarity metric, we employed rank distance (RD) based on function words frequencies - which was successfully used in previous studies concerning authorship problems (Popescu and Dinu, 2008; Dinu et al., 2008). The method is further detailed in Section 2.2.

The second approach was the Unmasking method proposed by Koppel et al. (Koppel et al., 2007), which evaluates how steeply the accuracy of a Support Vector Machine (SVM) model declines as the most informative features that differentiate two authors are gradually removed. The method is described in detail in Section 2.3.

2.1 Data

Our corpus is composed of 30 Italian authors, each being represented by at least four works. We included 14 novels written by Domenico Starnone and 8 of Elena Ferrante's novels. In order to test the Anita Raja hypothesis, we included 8 novels translated by Raja into Italian, as well as excerpts from interviews, prefaces and afterwords authored by her. The complete description of the corpus can be found in the Appendix (Table 1).

We preprocessed all texts by removing punctuation, lowercasing all letters and deriving the lemmas using the TreeTagger POS tagger (Schmid, 1995).

2.2 Clustering via Rank distance

We start from the assumption that, in order to assess the so-called stylome of an individual writer, we need to identify some parameters that are present across all texts regardless of its content. Thus, these parameters need to be unambiguous, quantifiable and used in an unconscious way. For precisely these reasons, function words are often used as stylistic indicators in computational analyses of texts across different languages (Chung and Pennebaker, 2011). Function words have successfully been employed in computational authorship attribution in cases such as *The Federalist Papers* in English, the case of the Mateiu Caragiale impersonators in Romanian (Dinu et al., 2008), etc. We chose a list of Italian function words composed of 134 words (Ranks.nl, n.d.) to which we added one more entry to compensate for the forms of the verb "essere" lost through lemmatization. The full list of functional words is presented in Fig. 8.

We then filtered out all the content words, reducing our text entries to function words which we ordered according to frequency in descending order, as a preparation step for calculating the Rank distances.

Rank distance (Dinu, 2003) is an ordinal metric related to the Spearman's footrule . In order to calculate the rank distance between two texts, the raw frequencies in the ordered lists are replaced with their ranks. The word with the highest frequency will be assigned the highest rank, while the words that are present on the function words list but absent in the text will be assigned the rank 0. Tied objects are assigned the average rank of the positions they share. As a result, we obtain a data frame that displays the rank of each function word within each specific text. We then compute the pairwise ranking distance between each two texts according to the following formula (L1 norm), where $\sigma(i)$ represents the rank of the object.

$$D(\sigma_1, \sigma_2) = \sum_{i=1}^{n} |\sigma_1(i) - \sigma_2(i)|$$

Since we calculate the pairwise rank distances between all texts, the result is a square similarity matrix, which serves as input for the Hierarchical Clustering Algorithm. For all our experiments we used Hierarchical Clustering Algorithm with average linkage. Previous studies showed that dendograms obtained through this method are highly accurate in grouping together texts written by the same authors (Popescu and Dinu, 2008; Dinu et al., 2008, 2012).

2.3 Unmasking

Our second approach was the Unmasking method (Koppel et al., 2007) which measures the rate of degradation of a classifier accuracy as features are iteratively removed. The underlying idea of this method is the fact that only a small number of features would contribute to the differentiation between texts pertaining to the same author. Thus, when we train a classifier to distinguish between texts written by the same author, we would observe a drastic decline of accuracy as the most distinguishing features are removed. We reproduced the method as described by Koppel in the following way, testing in turns one book against a corpus of several books written by a single author:

1. Define X as the book we want to analyze and A as a collection of books written by a single author.

2. Break all texts into chunks of 500 words each.

3. Identify the 250 most frequent words across X and A to use as features.

4. Represent each chunk as a numerical feature vector using the Bag of Words (BoW) approach.

5. Balance the datasets by randomly discarding chunks from A so that X and A contain the same number of chunks.

   (a) Repeat 10 times:

   (b) Train a linear-kernel Support Vector Machine (SVM) to distinguish between X and A using 10-fold cross-validation.

   (c) Find the top 3 negative and top 3 positive distinguishing features and remove them.

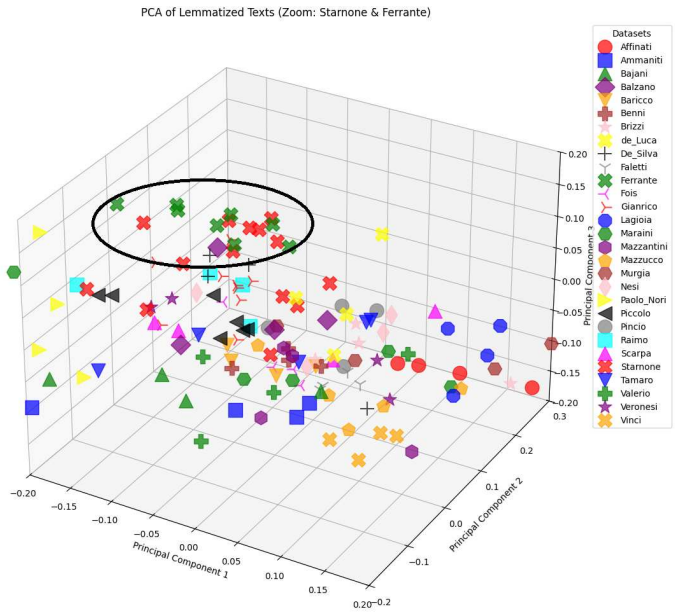


Figure 1: 3D PCA visualization of novels

6. Repeat the entire process (Step 5) 10 times, each time with a different random selection of text chunks.

7. Plot the average accuracy at each step to observe how the classifier performs as distinguishing words are removed.

3 Experiments and Analysis

We will now discuss in detail the results of our experiments.

3.1 Exploratory Analysis - PCA Visualization

The first step we performed was Exploratory Data Analysis (EDA). We chose to perform a 3D PCA visualization on the full corpus in order to gain some intuition about the similarities of the novels in our corpus. As mentioned above, we pre-processed the texts by removing punctuation and lemmatizing the words. For this first preliminary step, we chose to retain all the words (function words and content words) in order to capture as much stylistic and lexical variation as possible. Subsequently, we performed Principal Component Analysis (PCA) on the TF-IDF representations of the texts. The results shown in Fig. 1 show a clustering tendency for Elena Ferrante (green "X" symbol) and Domenico Starnone's novels representations (red "X" symbol).

3.2 Clustering via Rank Distance

We started the experiments with the Clustering via RD method - a method which aims to capture the similarity among texts according to the similarity of the ranks of the functional words pertaining to each text. By focusing solely the functional words, the method is able to retain the actual *stylistic fingerprint* of each author, rather than lexical differences that might arise from the varying topics of the novels. As per the methodology detailed in Section 2, we calculated the pairwise RD of the functional words frequencies extracted from each text, and the resulting square matrix was used as input for the Hierarchical Clutering model.

3.2.1 Rank Distance Experiment 1: All Authors

To check whether the RD Clustering method accurately clusters together texts written by the same authors, in the first experiment we included all works available in our corpus(Fig. 2). We observe that texts written by the same authors are clustered together with high precision. In the case of Ferrante (red) and Starnone (blue) on the other hand, not only are they clustered next to one another, but there is no demarcation between the two authors. The exception that we are going to encounter throughout all our experiments is Starnone's *Ex cattedra*, his first work which in our experiments shows little to no resemblance to his other works. This is consistent with previous research employing different methods (Tuzzi et al., 2018a). One possible explanation is the literary style of the novel, which takes the form of a diary. As we will see in the experiments involving non-literary works, the clustering method is highly sensible to variations in texts categories (literary vs. non-literary). To test the quality of the clusters, we calculated and obtained Cophenetic Correlation Coefficient of 0.78. This suggests that the clustering structure is reasonably well-preserved, as CCC measures how faithfully the hierarchical clustering represents the original pairwise distances (Sokal and Rohlf, 1962). This first experiment points towards Domenico Starnone as the very likely author behind Ferrante's texts, strengthening the conclusion reached by previous research. In order to further this claim, we conducted several additional experiments.

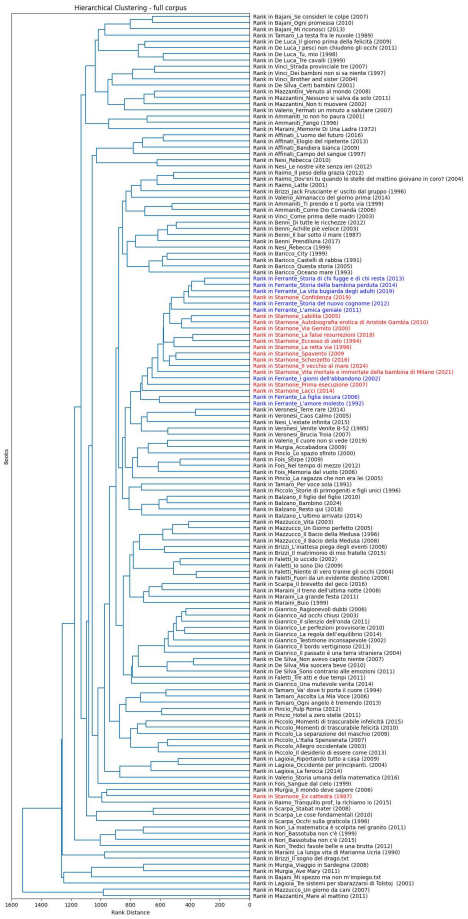


Figure 2: Full Corpus Clustering

3.2.2 Rank Distance Experiment 2: Just One Novel

In our second experiment, we wanted to see how Elena Ferrante and Domenico Starnone's novels will be clustered when we only include one author's full works, along with just a single novel from the other. Specifically, we wanted to see whether Ferrante's novels are clustered alongside Starnone's because they share group similarities, or whether only certain Ferrante novels share similarities with Starnone.

- **Experiment 2.1: Iterative clustering 1**

In the first experiment of this type, the corpus contained all of Domenico Starnone's novels and, iteratively, only one of Ferrante's works. Without any exceptions, each of Ferrante's novels was clustered next to Starnone's. Fig. 3 ilustrates one of the clusters obtained, while the others can be found in the Appendix (Fig. 9-16).

- **Experiment 2.2: Iterative clustering 2**

In order to test whether the reverse stands true as well, we repeated the experiments by keeping the

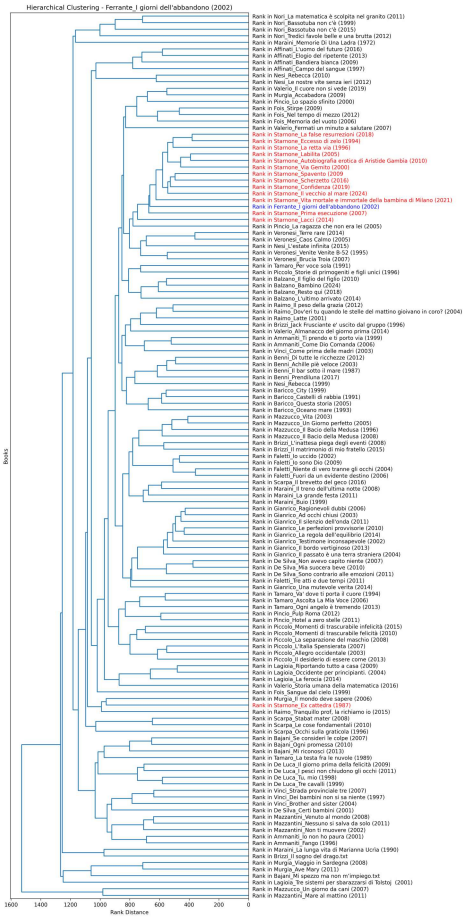


Figure 3: I giorni dell'abbandono vs. Full Corpus

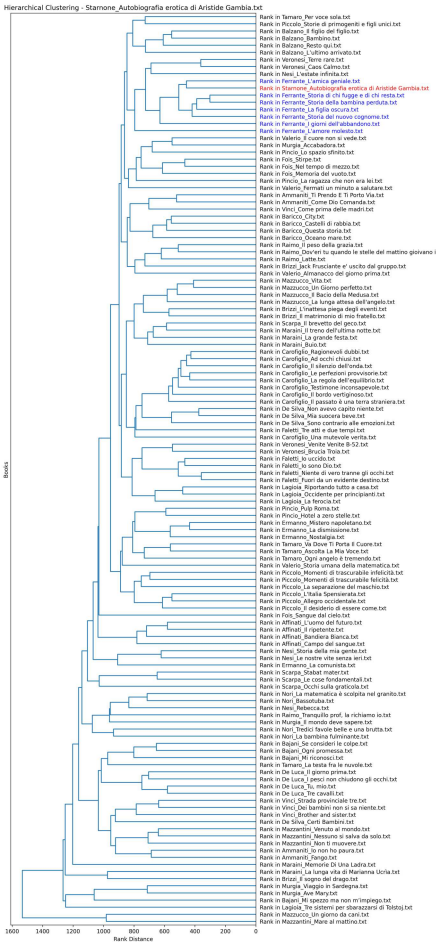


Figure 4: Autobiografia erotica di Aristide Gambia vs. Full Corpus

full Ferrante corpus in the clustering data frame and iteratively keeping only one of Starnone's novels. As seen in Fig. 4, the results mirrored the previous findings and each of Starnone's novels clustered alongside the Ferrante. For the full experiment results, see Appendix, Fig. 17-30.

3.2.3 Rank Distance Experiment 3: No Starnone

Finally, we removed all of Starnone's works from the corpus to test whether iteratively clustering a single novel by Ferrante would result in a close match with another author. As anticipated, when only one of Ferrante's novels was being clustered, each iteration grouped it with a different writer. This outcome confirmed that no other author in the corpus shares the same level of similarity with all of Ferrante's novels as is the case of Starnone. The full experiment results are presented in the Appendix, Fig. 31-38.

3.2.4 Rank Distance Experiment 4: Anita Raja

Having confirmed that Domenico Starnone is a strong candidate as the writer behind Ferrante's novels, we moved on to testing the Anita Raja hypothesis. Since Anita Raja is not a fiction writer herself, we had to rely on fiction books translated by Anita Raja and on non-fiction writings such as prefaces and afterwords. Thus, we expanded the corpus with the texts undisputedly authored by Anita Raja: 8 novels translated by Anita Raja and 7 prefaces and afterwords authored by her. We ran the experiment on the whole corpus and, according to our metric, there is no evident similarity between Elena Ferrante's and Raja's (green). When we removed Domenico Starnone from the corpus but kept all of Ferrante's works, Elena Ferrante's novels were clustered in the vicinity of Raja's translations, but in a clearly separate cluster.

Finally, when we iteratively ran the experiment on Ferrante's novels in the absence of Domenico

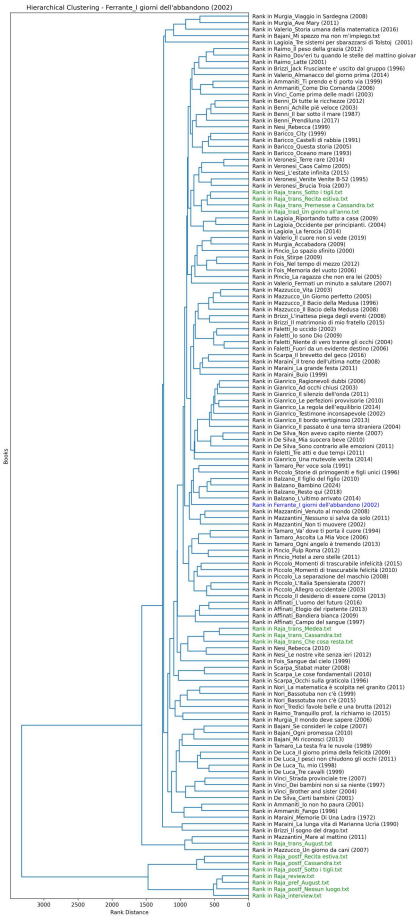


Figure 5: I giorni dell'abbandono vs. Raja Corpus

Starnone, only three of Ferrante's novels were clustered in the vicinity of Raja (eg. Fig. 5).

All of these results suggest that Raja is unlikely to be the author behind Elena Ferrante.

We note the fact that Raja's texts were separated into two different clusters: the translations and the non-literary texts. The non-literary texts cluster was completely separated from the rest of the clusters, which suggests that style homogenity (in the sense all texts should pertain to the same text category, be it literary or non-literary) is important for the method's accuracy. It is possible that the RD dis-similarity is caused by the fact that we compared a novel with a non-literary text. Thus, we acknowledge that a different corpus, tailored to non-fiction writings, could yield more definitive results.

3.3 Unmasking

To further test the Starnone hypothesis with a separate approach, we chose the unmasking method developed by Koppel and Schler. One of the advantages of this approach is that it doesn't necessarily require that the real author of the analyzed text is among the proposed candidates. In other words, if none of the proposed candidates authored Ferrante's novels, none of the accuracy plots will show an abrupt decrease as discriminative features are removed. We ran several texts with different chunks sizes (250, 500 and 1000 words) and features (stopwords only and all words). The results were highly similar and all pointed outlined Domenico Starnone in stark contrast with the rest of the candidates. For space reasons, we will only report the results obtained with the following parameters: all words and chunks size of 500 words.

3.3.1 Unmasking Elena Ferrante

To better illustrate the results of the unmasking experiments, we will present our results in 3 iterative steps, discussing the results we obtained for Elena Ferrante's novel *I giorni dell' abbandono*. The results obtained for the rest of the novels in the corpus can be found in the Appendix (Fig. 39). We tested Elena Ferrante's novel against 3 versions of the initial corpus. Importantly, whenever Ferrante was included in the comparison corpus, the novel currently being analyzed itself was excluded from the Ferrante corpus. The three corpus configurations were as follows:

- Corpus 1: Excluding both Domenico Starnone and Elena Ferrante
- Corpus 2: Excluding only Domenico Starnone
- Corpus 3: Full corpus, including both Domenico Starnone and Ferrante

In the first experiment (Fig. 6(a)) we applied the unmasking method on Elena Ferrante's novel against a corpus that included neither Domenico Starnone, nor other novels authored by Elena Ferrante. As mentioned in the Methodology section, according to the unmasking method the true author of a given text is the one for which we can observe a steep degradation in classification accuracy as discriminatory features are removed. As seen in the first plot that resulted from applying this method, the classification accuracy remained relatively high (around 0.7) as features were iteratively removed. This suggests that the differences between Elena Ferrante's novel and the novels authored by all the authors included in Corpus 1 run deep and remain consistent enough to accurately differentiate the authors even when being subjected to iterative removals.

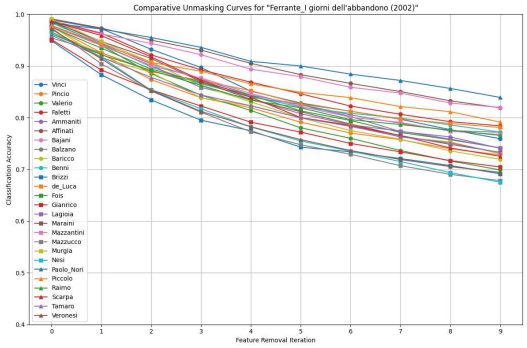
In the second experiment (Fig. 6(b)), we added the rest of Elena Ferrante's novels to the corpus (excluding *I giorno dell' abbandono*, the novel being tested). There is now a clear difference in the *steepness* of the accuracy drop as features are eliminated, especially in the first removal iterations. This suggests that all of Elena Ferrante's novels are written by the same author, and that the differences between the novels are rather shallow and due to topic differences, not to underlying unconscious style. When we remove only the top 6 differentiating features, there is a steep decline in the algorithm's ability to differentiate Elena Ferrante's novel *I giorni dell' abbandono* and the corpus comprised of the rest of Elena Ferrante's novels.

For the third experiment, we added Domenico Starnone's novels to the previous corpus. The resulting plot (Fig. 6(c)) confirmed our previous results: there is a great degree of similarity among Elena Ferrante's novels and those authored by Domenico Starnone. When trying to classifiy chunks of Elena Ferrante's novel *I giorni dell' abbandono* against chunks extracted from the corpus of novels written by Domenico Starnone using the Unmasking method, there is a very steep decrease in accuracy as the top differentiating features are removed - which suggests that the differences between the texts are shallow, probably caused by plot differences. Furthermore, the accuracy degradation curve resulting from Domenico Starnone's corpus (purple) is very similar to the one that resulted when the method was applied on the Elena Ferrante corpus (orange).

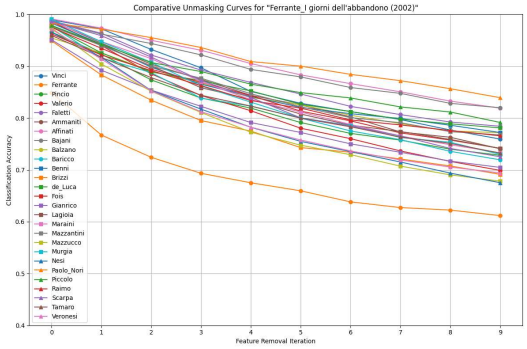
Similarly, for all the 8 of Elena Ferrante's novels, the unmasking method revealed Domenico Starnone as a clear candidate for the authorship of Elena Ferrante's novels. All the 8 plots that resulted in this experiment can be found in the Appendix (Fig. 39).

3.3.2 Unmasking Domenico Starnone

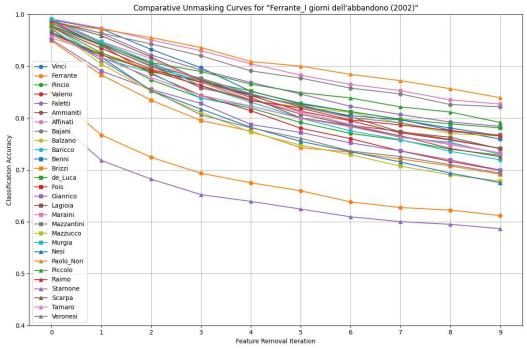
Once established that the unmasking method applied on the Ferrante corpus reveals Domenico Starnone as a possible author, we wanted to further explore the similarity of the novels signed by Elena Ferrante and Domenico Starnone by applying the unmasking algorithm on the Starnone corpus. Our hypothesis was that the reverse should also be true: if Domenico Starnone and Elena Ferrante are one and the same author, then by iteratively removing distinctive features when classifying chunks from a Starnone novel against the Ferrante corpus,



(a) *I giorni dell'abbandono* vs. Corpus 1



(b) *I giorni dell'abbandono* vs. Corpus 2



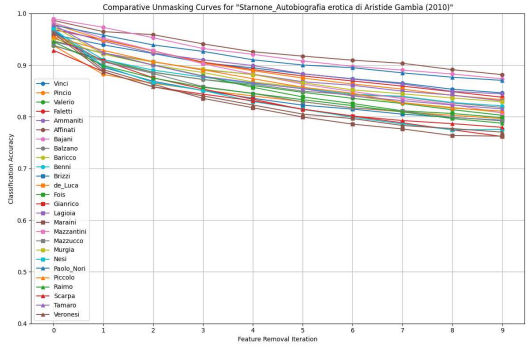
(c) *I giorni dell'abbandono* vs. Corpus 3

Figure 6: Unmasking Elena Ferrante

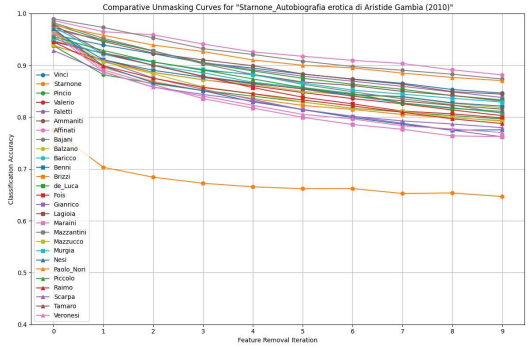
we would see a more pronounced decline in the accuracy for the Starnone-Ferrante pair than for Starnone against other authorship candidates.

To better illustrate our results, we applied the same 3-corpora method that we previously discussed in the case of Elena Ferrante: Corpus 1 Excluding both Domenico Starnone and Elena Ferrante, Corpus 2: Excluding only Elena Ferrante, Corpus 3 Full corpus, including both Domenico Starnone and Ferrante.

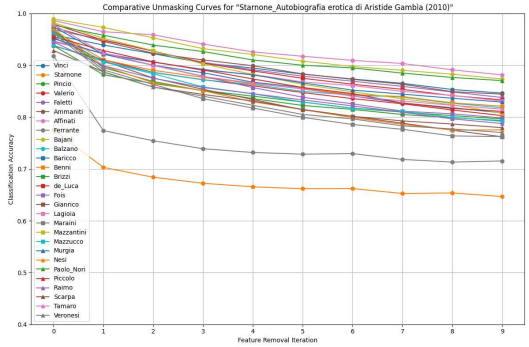
Our experimens have proven our hypothesis correct and mirror the results we had obtained for Elena Starnone. Fig. 7 shows the accuracy plot for Domenico's Starnone novel *Autobiografia erotica di Aristide Gambía* against the corpora of nov-



(a) *Autobiografia erotica di Aristide Gambía* vs. Corpus 1



(b) *Autobiografia erotica di Aristide Gambía* vs. Corpus 2



(c) *Autobiografia erotica di Aristide Gambía* vs. Corpus 3

Figure 7: Unmasking Domenico Starnone

els written by the other candidates. The sharpest decline in accuracy can be seen when Starnone's novel (orange) was classified against chunks extracted from the Ferrante corpus (grey). Thus, it is again proven that the differences between texts authored by Domenico Starnone and Elena Ferrante are shallow, and that by removing only 6 of the 250 features the accuracy decreases drastically. The same results were obtained for the full Starnone corpus with the exception of the 1987 novel *Ex cattedra* (See Appendix, Fig. 40).

Applying the unmasking method on Elena Ferrante and Domenico Starnone's novel reveals once again the high degree of similarity among the two authors. If we applied the unmasking method to any of the novels authored by Elena Ferrante in the absence of other texts written by her, it would appear as though Domenico Starnone was a sole clear candidate for the authorship - and the same situation is mirrored in the case of Domenico Starnone's novels when compared to texts authored by Elena Ferrante.

4 Conclusions

We investigated the complicated and fascinating authorship attribution case of Elena Ferrante in an attempt to test the Starnone hypothesis, who again and again is deemed to be the true author of the Ferrante novels by stylometric research, and assess the claims regarding Anita Raja - as so far only one computational study has taken her writings into account (Cortelazzo et al., 2024).

By using Clustering via Rank Distance and Unmasking as our two main authorship attribution methods, we aimed to cover a wide range of stylistic and methodological perspectives in order to increase the degree of confidence in our results.

Thus, we employed both supervised (Unmasking based on SVM models) and unsupervised (Clustering)learning methods. In addition, we ran experiments based on both functional words only (in the Clustering experiments) and the full vocabulary employed by the authors (in the Unmasking experiments, as studies such as (Juola, 2017) show that lexical choice expressed through content words may also contribute to building the lexical fingerprint of an author).

All of the clustering experiments revealed a great degree of similarity between Elena Ferrante and Domenico Starnone's works, while the Unmasking experiments showed that the features which discriminate between the two authors are shallow.

Although a wider non-fiction corpus that includes samples from different authors might help bring further and more definite proof, our results on the current corpus show that Anita Raja is not the author behind Elena Ferrante and that the profile of Domenico Starnone is the one that most resembles that of Elena Ferrante. Our findings align with previous research and bring further insights into the tight similarity between Ferrante and Starnone's works.

5 Limitations and Ethical Concerns

One of the main limitations of our study is the assumption that the real author must be among the proposed candidates. In reality, assembling a corpus that comprehensively includes all possible authors would be impossible. While we ensured that each author was represented by a minimum of four works, a more extensive corpus could have yielded more definitive results.

Secondly, in the case of Anita Raja, our analysis compares her non-fiction writings with several novels, meaning that some of the observed differences may be caused by genre variation rather than authorship alone. Future research should incorporate a corpus specifically designed for non-fiction works.

Finally, we acknowledge the ethical complexities surrounding our study, particularly in relation to an author who is still alive and has deliberately chosen to remain anonymous. Respecting an author's right to privacy must be carefully balanced with literary and academic inquiry, and our study is only driven by research motives and all our conclusions are tied to the stylometric similarities of written texts voluntarily published by the authors.

References

Cindy Chung and James Pennebaker. 2011. The psychological functions of function words. In *Social communication*, pages 343–359. Psychology Press.

Michele A Cortelazzo, George K Mikros, and Arjuna Tuzzi. 2024. Applying general impostors method to the ferrante case. *Digital Stylistics in Romance Studies and Beyond*, pages 299–313.

Hugh Craig and Arthur F Kinney. 2009. *Shakespeare, computers, and the mystery of authorship*. Cambridge University Press.

Liviu P Dinu. 2003. On the classification and aggregation of hierarchies with ifferent constitutive elements. *Fundamenta Informaticae*, 55(1):39–50.

Liviu P. Dinu, Vlad Niculae, and Octavia-Maria Șulea. 2012. Pastiche detection based on stopword rankings: Exposing impersonators of a romanian writer. In *Proceedings of the Workshop on Computational Approaches to Deception Detection, co-located with EACL 2012*, pages 72–77, Avignon, France. Held on April 23, 2012.

Liviu Petrisor Dinu, Marius Popescu, and Anca Dinu. 2008. Authorship identification of romanian texts with controversial paternity. In *Proceedings of the International Conference on Language Resources and Evaluation, LREC 2008, 26 May - 1 June 2008, Marrakech, Morocco*. European Language Resources Association.

Claudio Gatti. 2016. Elena ferrante: tracce dell'autrice ritrovata. *Il Sole 24 Ore*. Accessed: 2025-05-03.

David I Holmes. 1994. Authorship attribution. *Computers and the Humanities*, 28:87–106.

Patrick Juola. 2008. Authorship attribution. *Foundations and Trends® in Information Retrieval*, 1(3):233–334.

Patrick Juola. 2015. The rowling case: a proposed standard analytic protocol for authorship questions. *Digital Scholarship in the Humanities*, 30(suppl_1):i100–i113.

Patrick Juola. 2017. Thesaurus-based semantic similarity judgments: A new approach to authorial similarity? In *A. Tuzzi, & MA Cortelazzo (2018), Drawing Elena Ferrante's Profile. Workshop Proceedings, Padova*, pages 47–59.

Moshe Koppel, Jonathan Schler, and Shlomo Argamon. 2009. Computational methods in authorship attribution. *Journal of the American Society for information Science and Technology*, 60(1):9–26.

Moshe Koppel, Jonathan Schler, and Elisheva Bonchek-Dokow. 2007. Measuring differentiability: Unmasking pseudonymous authors. *Journal of Machine Learning Research*, 8(45):1261–1276.

Cyril Labbé and Dominique Labbé. 2001. Inter-Textual Distance and Authorship Attribution. Corneille and Molière. *Journal of Quantitative Linguistics*, 8(3):213–231. Version anglaise préliminaire à l'article paru sous ce titre dans le Journal of Quantitative Linguistics. 8-3, December 2001, p 213-231.

Frederick Mosteller and David L. Wallace. 1964. *Inference and Disputed Authorship: The Federalist*. Addison-Wesley.

Marius Popescu and Liviu Petrisor Dinu. 2008. Rank distance as a stylistic similarity. In *COLING 2008, 22nd International Conference on Computational Linguistics, Posters Proceedings, 18-22 August 2008, Manchester, UK*, pages 91–94.

Ranks.nl. n.d. Stopwords for italian. `https://www.ranks.nl/stopwords/italian`. Accessed: 2025-02-16.

Jan Rybicki. 2018. Partners in life, partners in crime. In Arjuna Tuzzi and Michele A. Cortelazzo, editors, *Drawing Elena Ferrante's Profile. Workshop Proceedings, Padova, 7 September 2017*, pages 111–122. Padova University Press, Padova.

Jacques Savoy. 2018a. Elena ferrante unmasked. In Arjuna Tuzzi and Michele A. Cortelazzo, editors, *Drawing Elena Ferrante's Profile. Workshop Proceedings, Padova, 7 September 2017*, pages 123–141. Padova University Press, Padova. 2018a.

Jacques Savoy. 2018b. Is starnone really the author behind ferrante? *Digital Scholarship in the Humanities*, 33(4):902–918.

Helmut Schmid. 1995. Treetagger. https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/. Accessed: 2025-02-16.

Dan Shen. 2017. The bloomsbury companion to stylistics.

Robert R Sokal and F James Rohlf. 1962. The comparison of dendrograms by objective methods. *Taxon*, pages 33–40.

Efstathios Stamatatos. 2009. A survey of modern authorship attribution methods. *Journal of the American Society for information Science and Technology*, 60(3):538–556.

The New York Times. 2024. The 100 best books of the 21st century. https://www.nytimes.com/interactive/2024/books/best-books-21st-century.html. Accessed: 2025-05-02.

Arjuna Tuzzi, Michele A Cortelazzo, et al. 2018a. *Drawing Elena Ferrante's Profile. Workshop Proceedings, Padova, 7 September 2017*. Padova University Press.

Arjuna Tuzzi, Michele A Cortelazzo, et al. 2018b. It takes many hands to draw elena ferrante's profile. *Drawing Elena Ferrante's Profile*, pages 9–30.

Hans Van Halteren, Harald Baayen, Fiona Tweedie, Marco Haverkort, and Anneke Neijt. 2005. New machine learning methods demonstrate the existence of a human stylome. *Journal of Quantitative Linguistics*, 12(1):65–77.

Figure 8: List of Italian Stopwords

a	adesso	ai	al	alla
allo	allora	altre	altri	altro
anche	ancora	avere	aveva	avevano
ben	buono	che	chi	cinque
comprare	con	consecutivi	consecutivo	cosa
cui	da	del	della	dello
dentro	deve	devo	di	doppio
due	e	ecco	fare	fine
fino	fra	gente	giu	ha
hai	hanno	ho	il	indietro
invece	io	la	lavoro	le
lei	lo	loro	lui	lungo
ma	me	meglio	molta	molti
molto	nei	nella	no	noi
nome	nostro	nove	nuovi	nuovo
o	oltre	ora	otto	peggio
pero	persone	piu	poco	primo
promesso	qua	quarto	quasi	quattro
quello	questo	qui	quindi	quinto
rispetto	sara	secondo	sei	sembra
sembrava	senza	sette	sia	siamo
siete	solo	sono	sopra	soprattutto
sotto	stati	stato	stesso	su
subito	sul	sulla	tanto	te
tempo	terzo	tra	tre	triplo
ultimo	un	una	uno	va
vai	voi	volte	vostro	

A Appendix

Author	**Works**
Eraldo Affinati	*Bandiera Bianca, Campo del sangue, Elogio del ripetente, L'uomo del futuro*
Niccolò Ammaniti	*Fango, Ti prendo e ti porto via, Io non ho paura, Come Dio comanda*
Andrea Bajani	*Mi spezzo ma non m'impiego, Se consideri le colpe, Ogni promessa, Mi riconosci*
Marco Balzano	*Il figlio del figlio, L'ultimo arrivato, Bambino, Resto qui*
Alessandro Baricco	*Castelli di rabbia, Oceano mare, City, Questa storia*
Stefano Benni	*Il bar sotto il mare, Di tutte le ricchezze, Prendiluna, Achille piè veloce*
Enrico Brizzi	*Jack Frusciante è uscito dal gruppo, L'inattesa piega degli eventi, Il matrimonio di mio fratello, Il sogno del drago*
Gianrico Carofiglio	*Testimone inconsapevole, Ad occhi chiusi, Il passato è una terra straniera, Ragionevoli dubbi, Le perfezioni provvisorie, Il silenzio dell'onda, Il bordo vertiginoso delle cose, Una mutevole verità, La regola dell'equilibrio*
Erri De Luca	*Tu, mio, Tre cavalli, Il giorno prima della felicità, I pesci non chiudono gli occhi*
Diego De Silva	*Certi bambini, Non avevo capito niente, Mia suocera beve, Sono contrario alle emozioni*
Giorgio Faletti	*Io uccido, Niente di vero tranne gli occhi, Fuori da un evidente destino, Io sono Dio, Tre atti e due tempi*
Elena Ferrante	*L'amore molesto, I giorni dell'abbandono, La figlia oscura, L'amica geniale, Storia del nuovo cognome, Storia di chi fugge e di chi resta, Storia della bambina perduta, La vita bugiarda degli adulti*
Marcello Fois	*Stirpe, Nel tempo di mezzo, Sangue dal cielo, Memoria del vuoto*
Nicola Lagioia	*Tre sistemi per sbarazzarsi di Tolstoj, Riportando tutto a casa, La ferocia, Occidente per principianti*
Dacia Maraini	*Memorie di una ladra, La lunga vita di Marianna Ucria, Buio, Il treno dell'ultima notte, La grande festa*
Margareth Mazzantini	*Non ti muovere, Venuto al mondo, Mare al mattino, Nessuno si salva da solo*

Author	Works
Melania G. Mazzucco	*Il bacio della Medusa, Vita, Un giorno perfetto, Un giorno da cani, La lunga attesa dell'angelo*
Michela Murgia	*Il mondo deve sapere. Romanzo tragicomico di una telefonista precaria, Viaggio in Sardegna. Undici percorsi nell'isola che non si vede, Accabadora, Ave Mary. E la Chiesa inventò la donna*
Edoardo Nesi	*Rebecca, Storia della mia gente. La rabbia e l'amore della mia vita di industriale di provincia, L'estate infinita, Le nostre vite senza ieri*
Paolo Nori	*Bassotuba non c'è, La matematica è scolpita nel granito, Tredici favole belle e una brutta, La bambina fulminante*
Francesco Piccolo	*Storie di primogeniti e figli unici, Allegro occidentale, L'Italia spensierata, La separazione del maschio, Momenti di trascurabile felicità, Il desiderio di essere come tutti, Momenti di trascurabile infelicità*
Tommaso Pincio	*Lo spazio sfinito, Hotel a zero stelle. Inferni e paradisi di uno scrittore senza fissa dimora, Pulp Roma, La ragazza che non era lei*
Christian Raimo	*Latte, Il peso della grazia, Dov'eri tu quando le stelle del mattino gioivano in coro?, Tranquillo prof, la richiamo io*
Tiziano Scarpa	*Occhi sulla graticola, Stabat Mater, Le cose fondamentali, Il brevetto del geco*
Domenico Starnone	*Ex cattedra, Eccesso di zelo, Via Gemito, Prima esecuzione, Autobiografia erotica di Aristide Gambìa, Lacci, Scherzetto, Confidenza, Il vecchio al mare, Le false resurrezione, Labilità, La retta via. Otto storie di obiettivi mancati, Spavento, Vita mortale e immortale della bambina di Milano*
Susanna Tamaro	*La testa tra le nuvole, Per voce sola, Va' dove ti porta il cuore, Ascolta la mia voce, Ogni angelo è tremendo*
Chiara Valerio	*Fermati un minuto a salutare, Almanacco del giorno prima, Storia umana della matematica, Il cuore non si vede*
Sandro Veronesi	*Venite venite B-52, Caos calmo, Brucia Troia, Terre rare*
Simona Vinci	*Dei bambini non si sa niente, Brother and Sister, Strada Provinciale Tre, Come Prima Delle Madri*

Table 1: Full Corpus

Figure 9: Experiment 2.1 - Ferrante, La figlia oscura

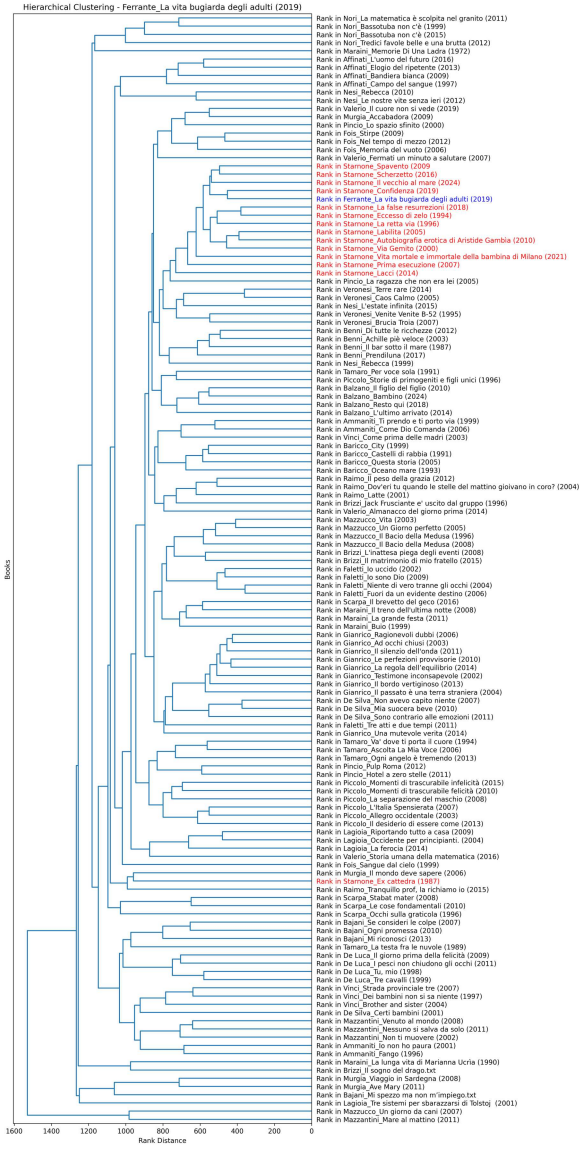


Figure 10: Experiment 2.1 - Ferrante, La vita bugiarda degli adulti

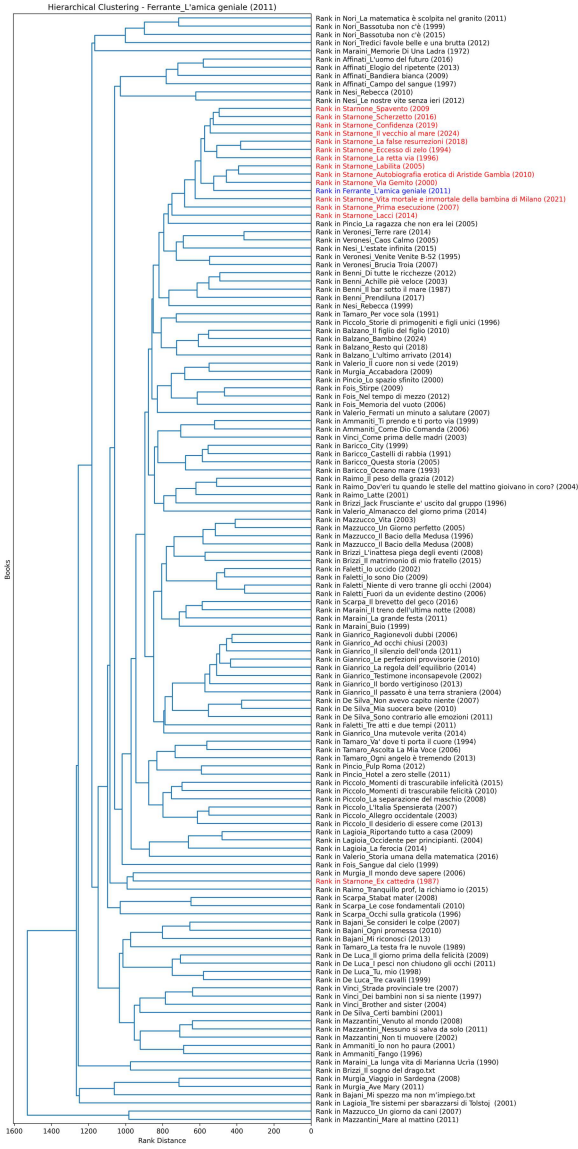


Figure 11: Experiment 2.1 - Ferrante, L'amica geniale

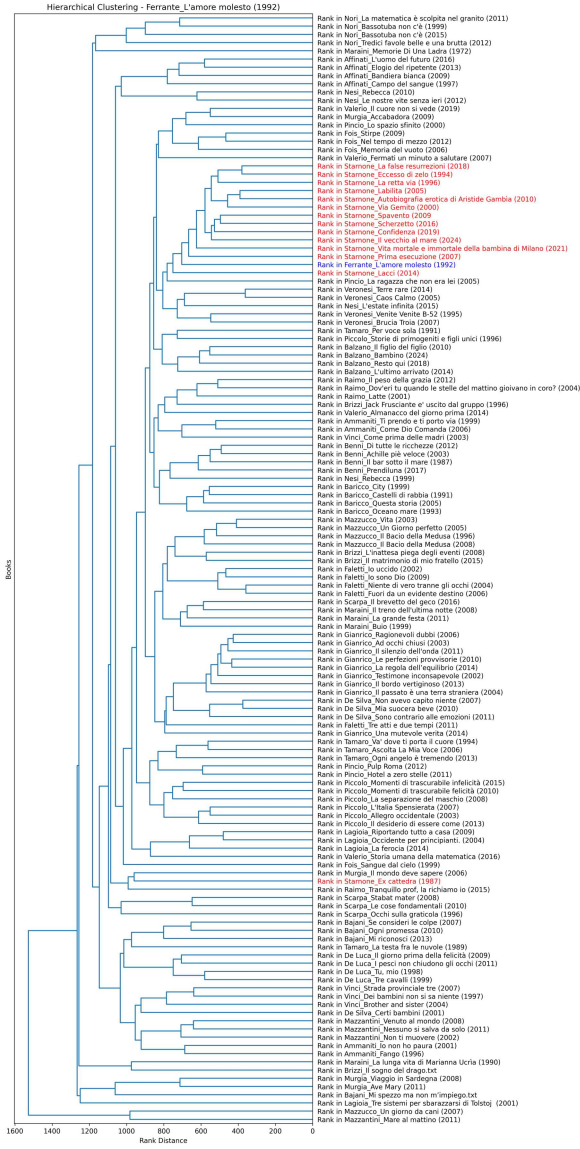


Figure 12: Experiment 2.1 - Ferrante, L'amore molesto

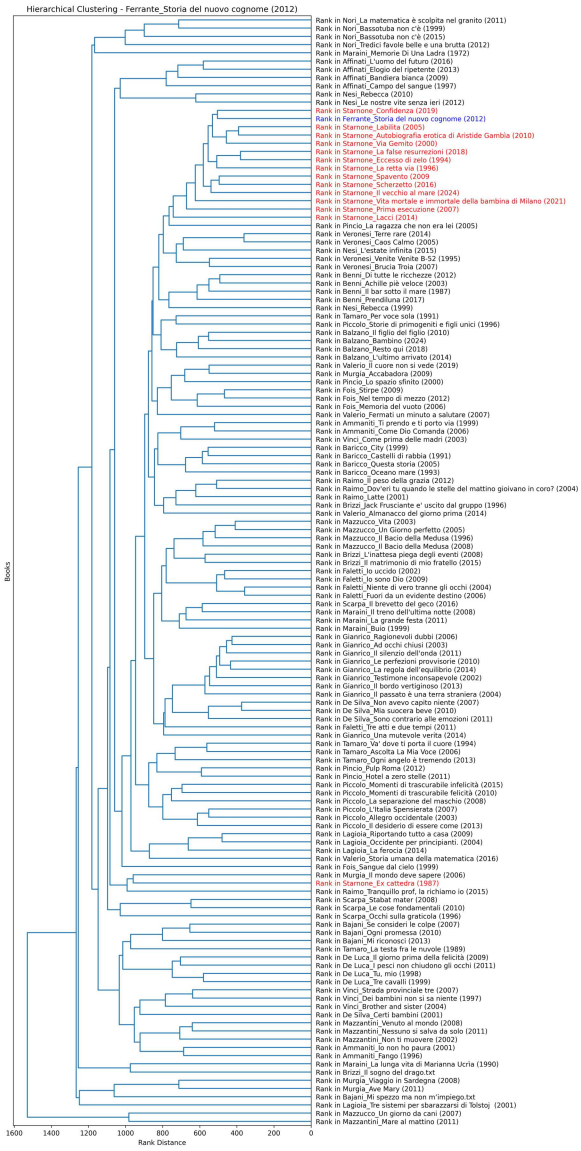


Figure 13: Experiment 2.1 - Ferrante, Storia del nuovo cognome

Figure 14: Experiment 2.1 - Ferrante, Storia della bambina perduta

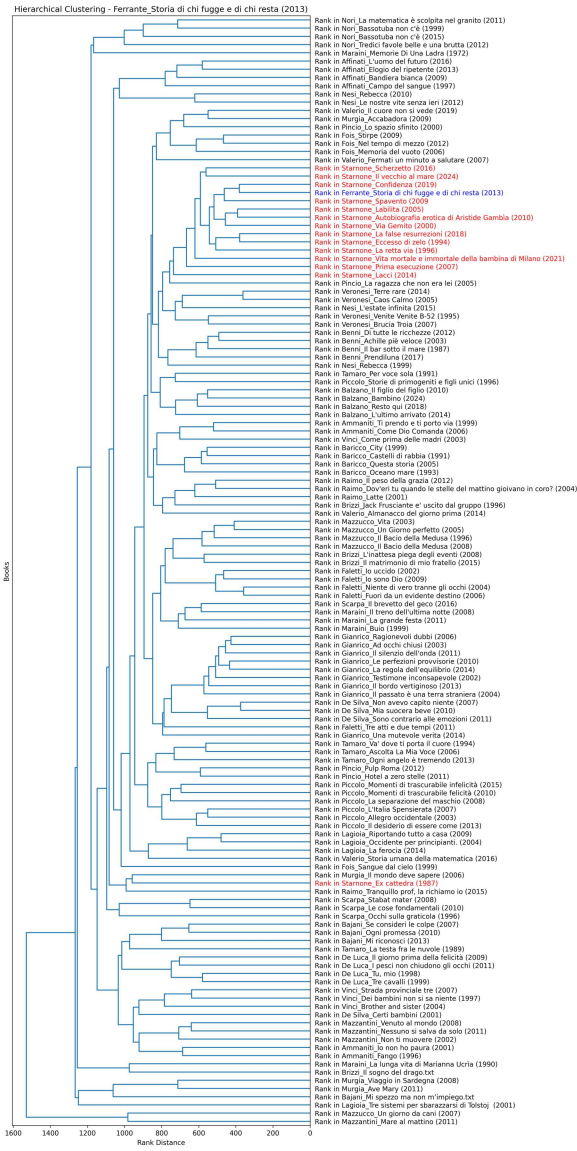


Figure 15: Experiment 2.1 - Ferrante, Storia di chi fugge e di chi resta

Figure 16: Experiment 2.2 - Starnone, Confidenza

Figure 17: Experiment 2.2 - Starnone, Eccesso di zelo

Figure 18: Experiment 2.2 - Starnone, Ex cattedra

Figure 19: Experiment 2.2 - Starnone, Il vecchio al mare

Figure 20: Experiment 2.2 - Starnone, La false resurrezioni

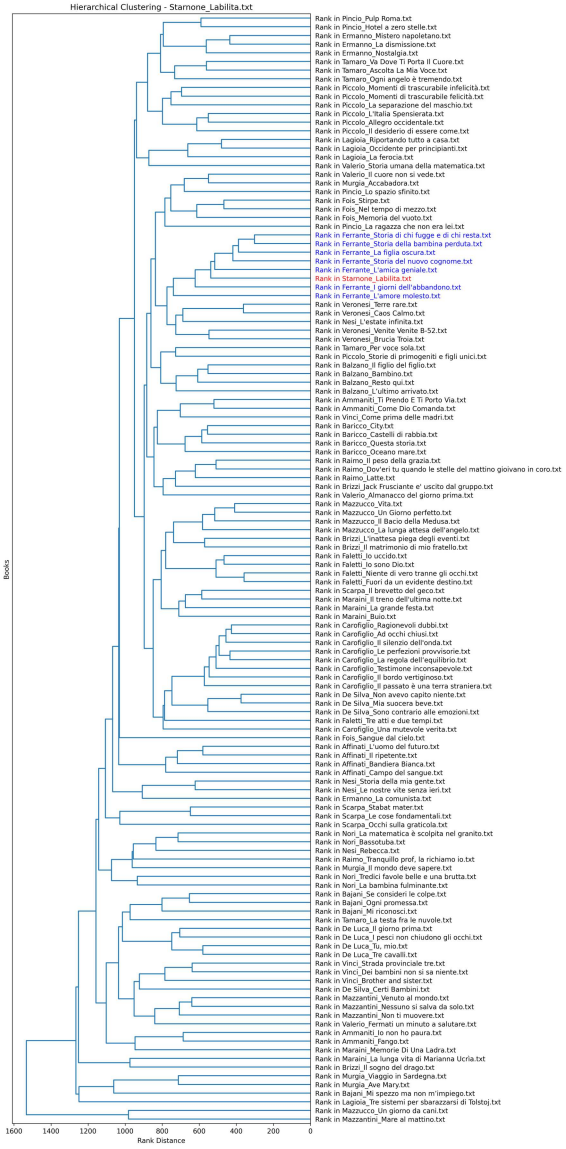


Figure 21: Experiment 2.2 - Starnone, Labilita

Figure 22: Experiment 2.2 - Starnone, Lacci

Figure 23: Experiment 2.2 - Starnone, Otto storie

Figure 24: Experiment 2.2 - Starnone, Prima esecuzione

Figure 25: Experiment 2.2 - Starnone, Scherzetto

Figure 26: Experiment 2.2 - Starnone, Spavento

Figure 27: Experiment 2.2 - Starnone, Via Gemito

Figure 28: Experiment 2.2 - Starnone, Vita mortale e immortale della bambina di Milano

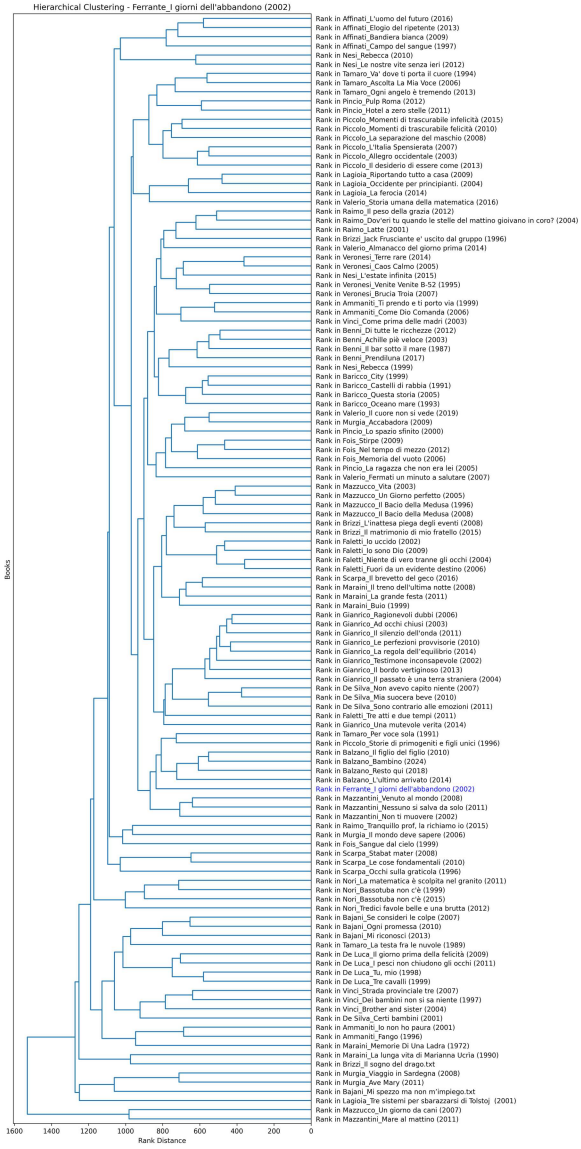


Figure 29: Experiment 3 - Ferrante, I giorni dell'abbandono

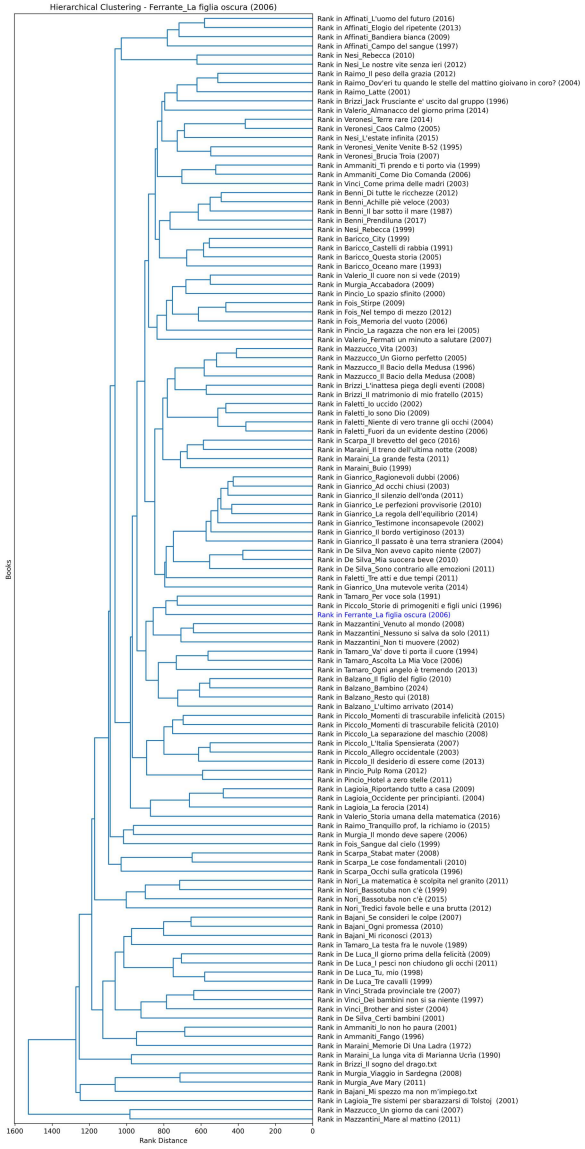


Figure 30: Experiment 3 - Ferrante, La figlia oscura

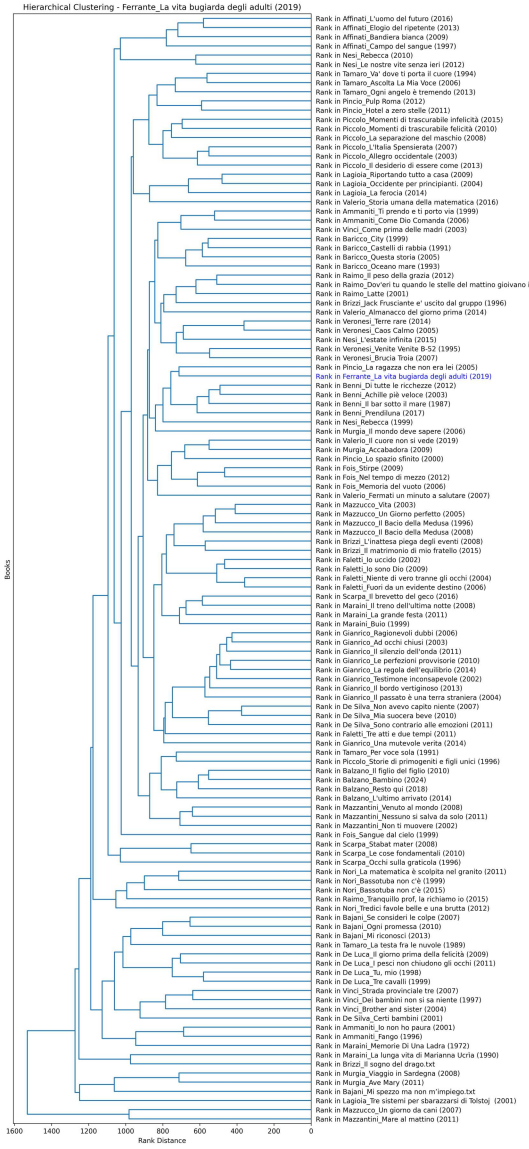


Figure 31: Experiment 3 - Ferrante, La vita bugiarda degli adulti

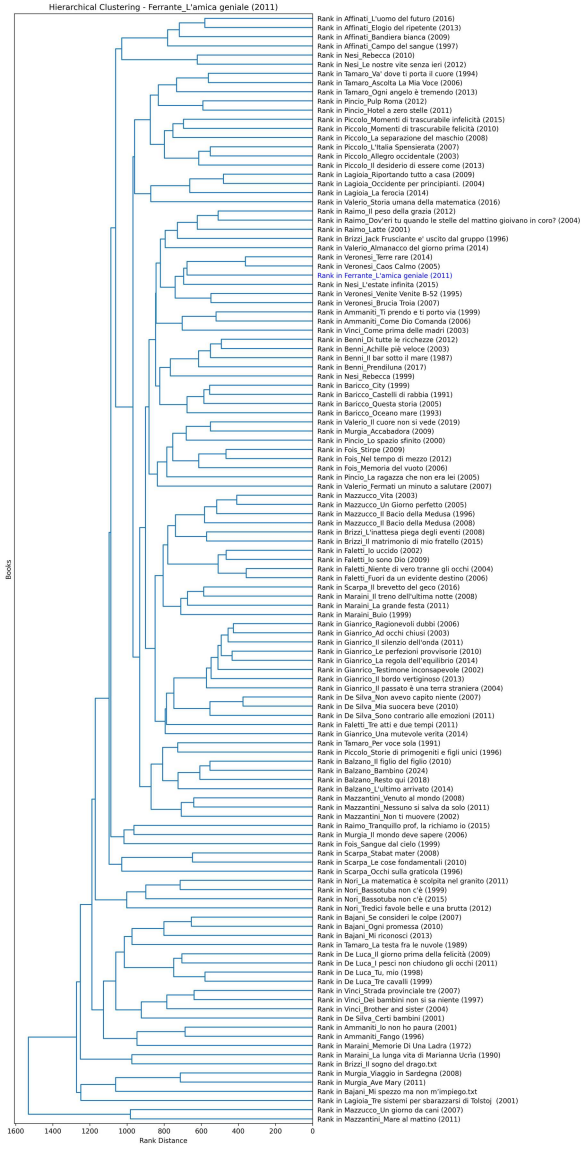


Figure 32: Experiment 3 - Ferrante, L'amica geniale

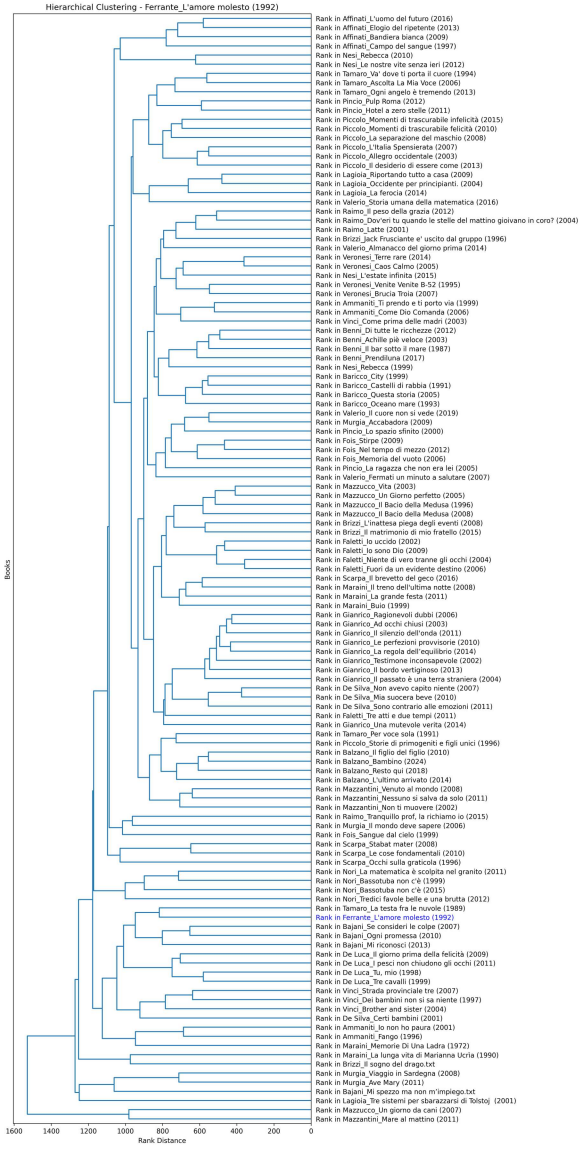


Figure 33: Experiment 3 - Ferrante, L'amore molesto

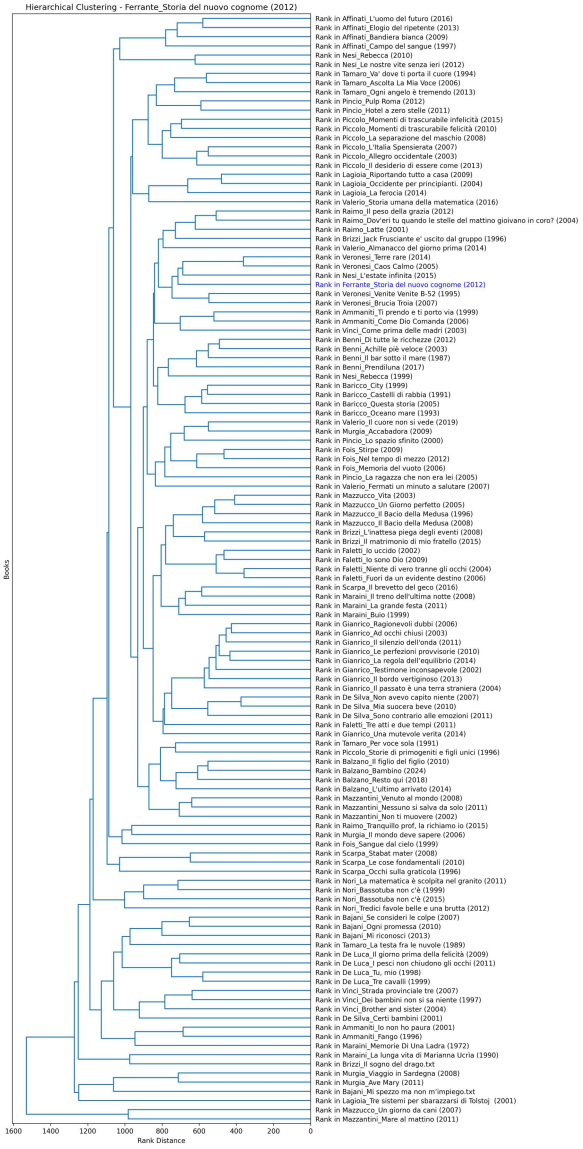


Figure 34: Experiment 3 - Ferrante, Storia del nuovo cognome

Figure 35: Experiment 3 - Ferrante, Storia della bambina perduta

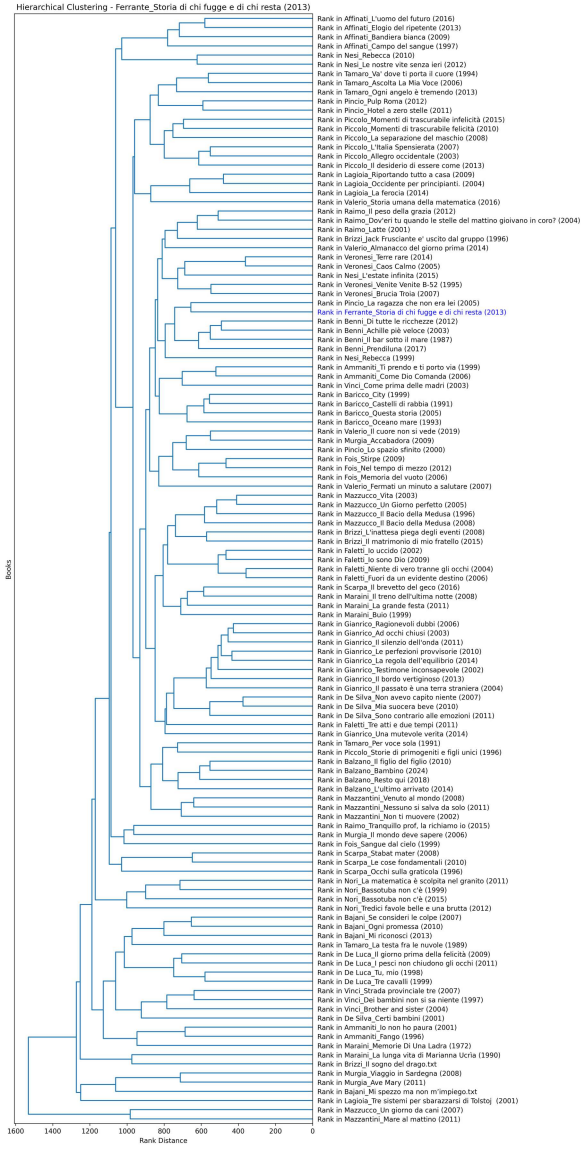
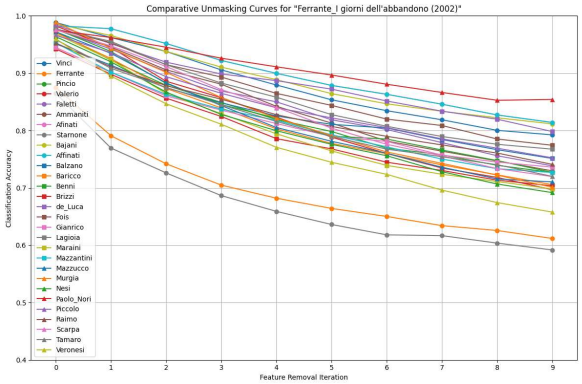
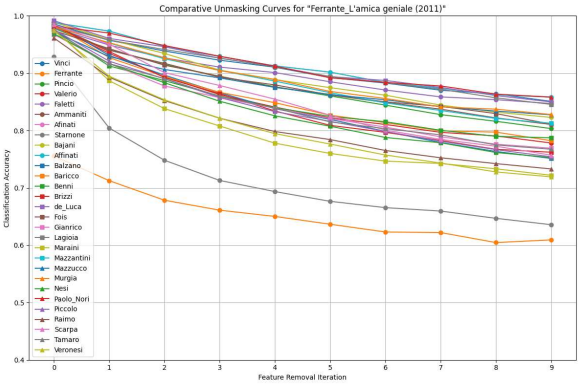


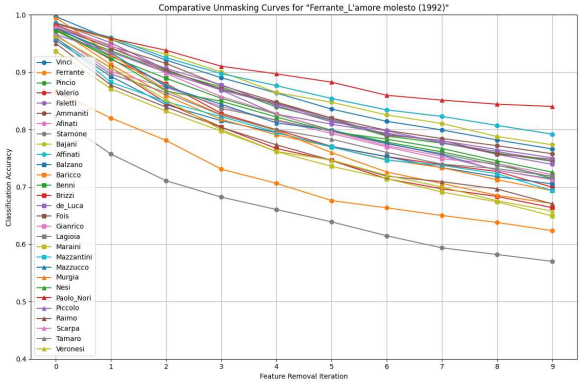
Figure 36: Experiment 3 - Ferrante, Storia di chi fugge e di chi resta



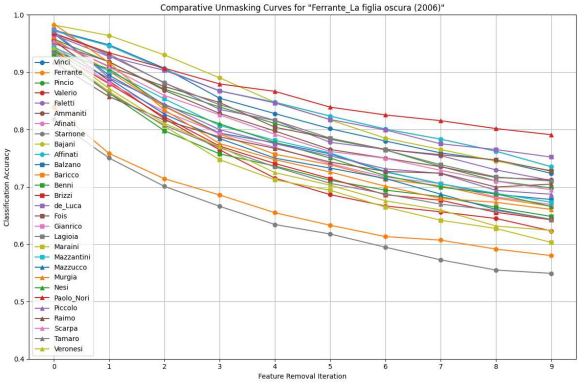
(a) Ferrante, I giorni dell'abbandono



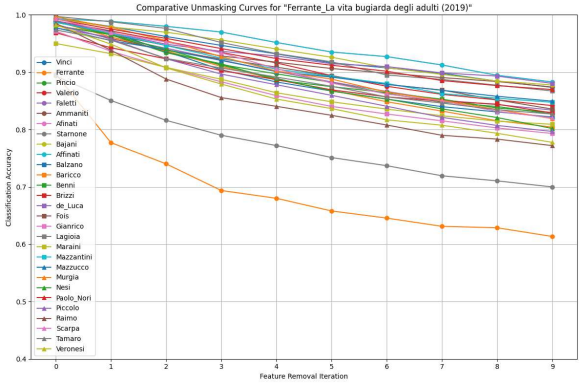
(b) Ferrante, La figlia oscura



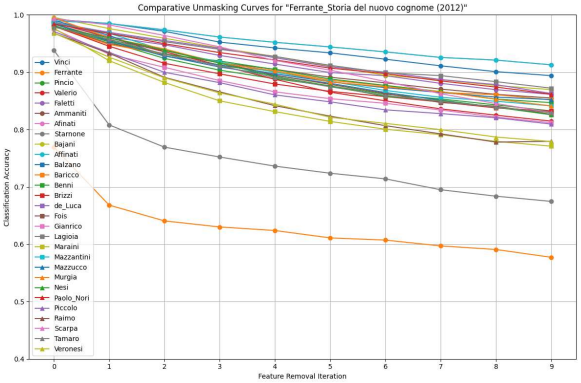
(c) Ferrante, La vita bugiarda degli adulti



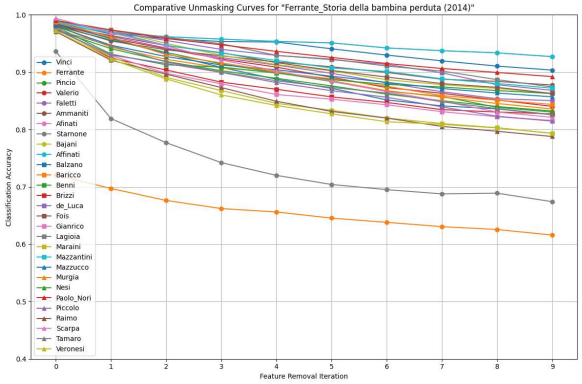
(d) Ferrante, L'amica geniale



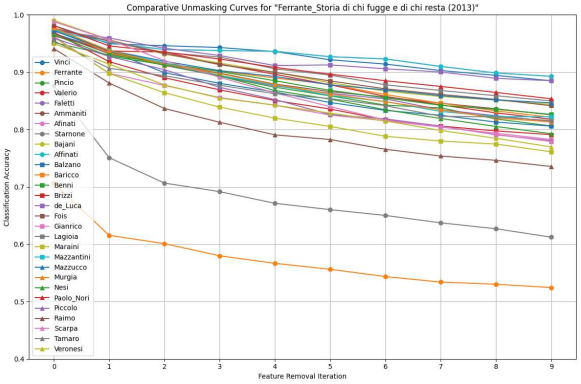
(e) Ferrante, L'amore molesto



(f) Ferrante, Storia del nuovo cognome

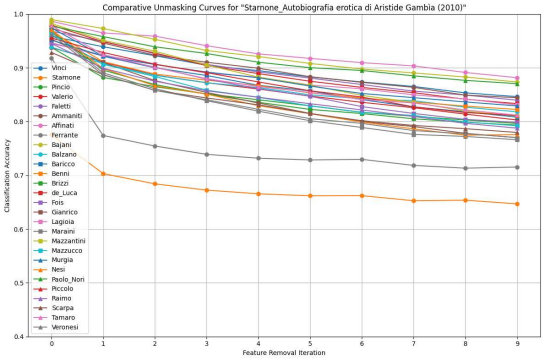


(g) Ferrante, Storia della bambina perduta

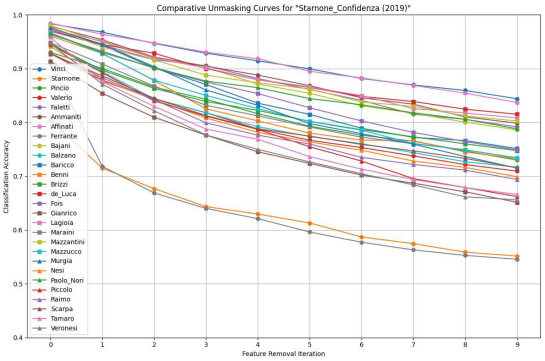


(h) Ferrante, Storia di chi fugge e di chi resta

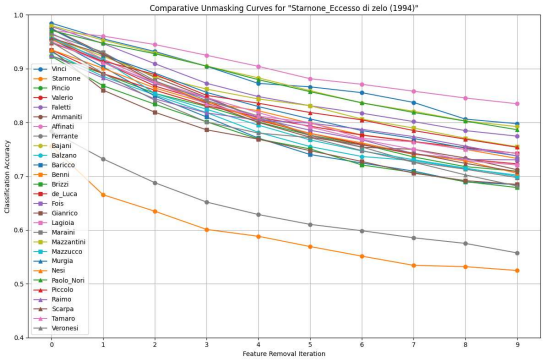
Figure 37: Unmasking Elena Ferrante



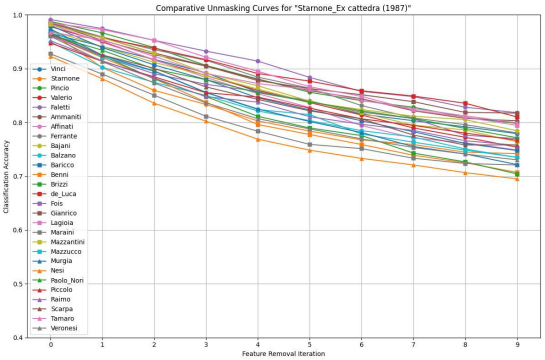
(a) Starnone, Autobiografia erotica di Aristide Gambia



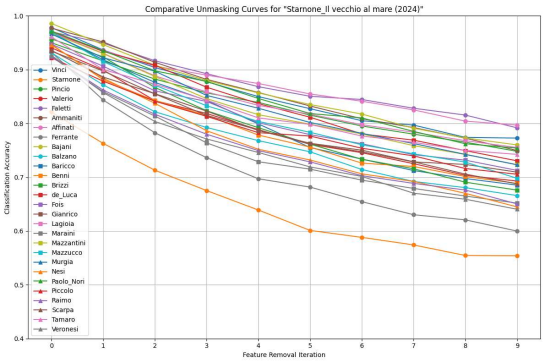
(b) Starnone, Confidenza



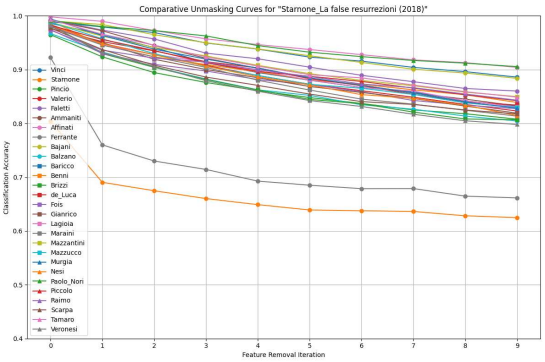
(c) Starnone, Eccesso di zelo



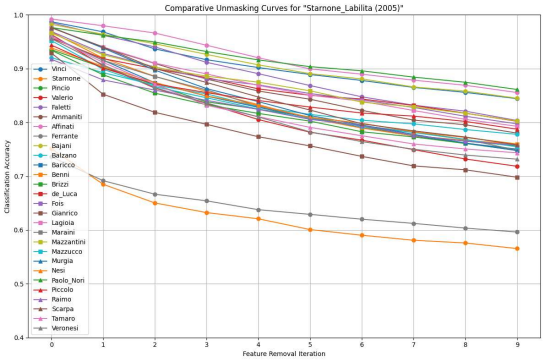
(d) Starnone, Ex cattedra



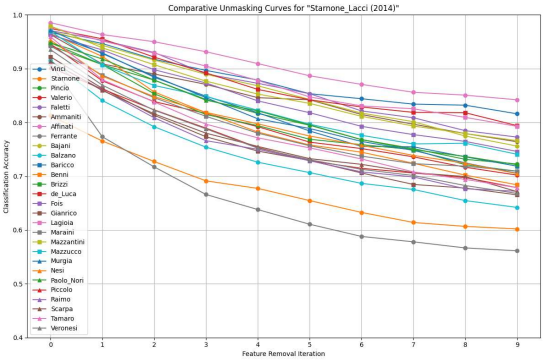
(e) Starnone, Il vecchio al mare



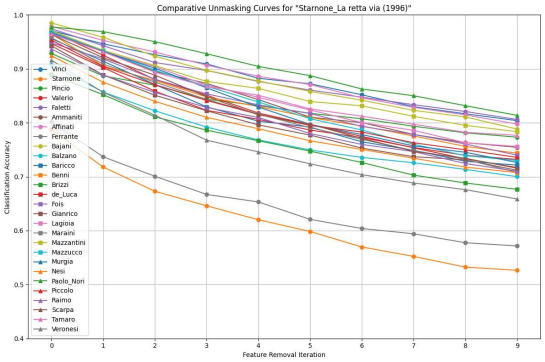
(f) Starnone, La false rezurrezioni



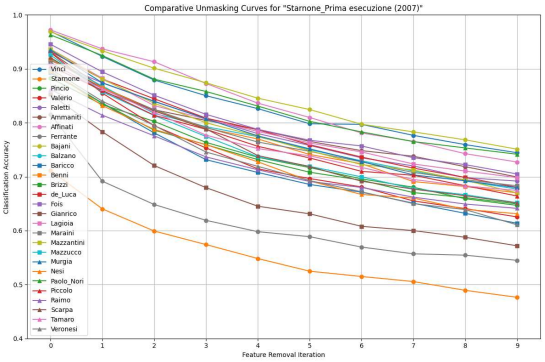
(g) Starnone, Labilità



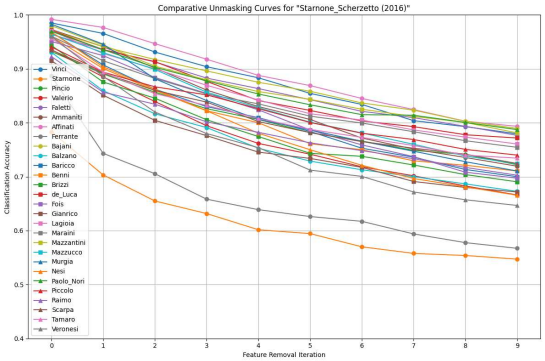
(h) Starnone, Lacci



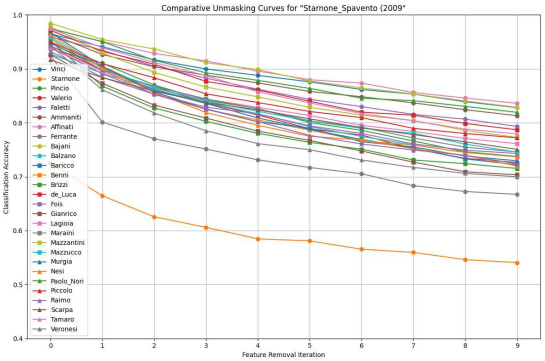
(i) Starnone, La retta via



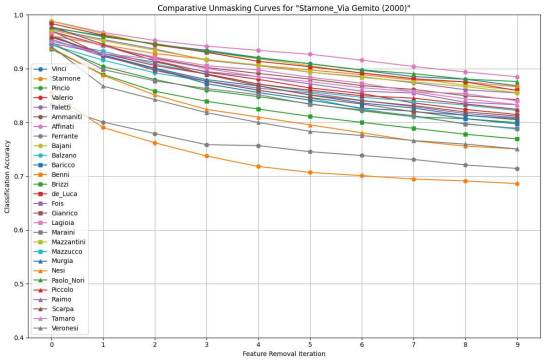
(j) Starnone, Prima esecuzione



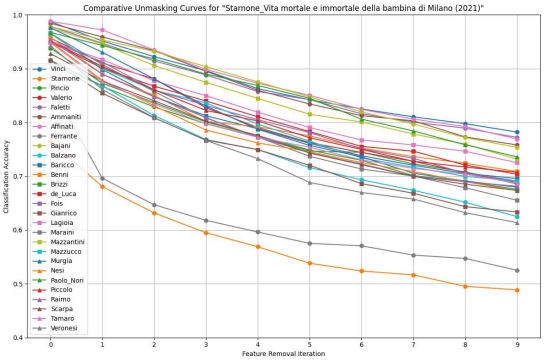
(k) Starnone, Scherzetto



(l) Starnone, Spavento



(m) Starnone, Via Gemito



(n) Starnone, Vita mortale e immortale della bambina di Milano

Figure 38: Unmasking Domenico Starnone